Overview of the bibliometric indicators of CWTS

appendix to the report 'Research Impact of the Dutch University Medical Centers'

Universiteit Leiden

Overview of bibliometric indicators

P	Number of papers (normal articles and reviews) published in journals processed for the Web of Science (WoS).
TCS	Number of citations recorded in WoS journals to all papers involved. Self-citations are excluded.
MCS	Average number of citations per publication, or citation per publication ratio. Self-citations are excluded.
MNCS	The impact of a research unit's articles, compared to the world citation average in the subfields in which the research unit is active.
MNJS	The impact of the journals in which a research unit has published (the research unit's journal selection), compared to the world citation average in the subfields covered by these journals.
PP(top 10%)	The share of the number of papers that are among the 10% most frequently cited of all similar papers in the period 2004-2015/20156

Explanation of the bibliometric methodology of CWTS

Universiteit Leiden

1. Data collection

1.1 Introduction

Bibliometrics is the quantitative study of written products of research. It is assumed that scientific subjects develop at an international research front (Price, 1963). Research results are communicated in publications that are submitted to evaluation by professional colleagues. In the references of their papers, scientists acknowledge relevant publications by others, as they build on previous work. Therefore, the number of times a publication is referred to gives a partial indication of the 'impact' of a publication, its reception and use by scientists at the research front.

In nearly all scientific fields, the scientific journal is an important medium of communication. The Web of Science database (from here on WoS), which consists of the citation indexes known under acronyms such as SCI, SSCI and A&HCI claims to cover the most important 'leading' international journals and serials (such as Annual Reviews) with a well-functioning referee system. In addition, the overall citation rate of journals is considered, as well as their timeliness of publication, and adherence to international editorial conventions. Regularly, a limited number of new journals are added, while other journals are no longer covered. More 'peripheral' journals, often national in scope, are usually not covered by the CI. The WoS counts about 11,000 journals during the last decade.

Both statistical requirements and imperfections in the citation make it desirable to aggregate across individuals, publications, and citations. As scientific (sub)fields differ in publication and citation patterns (as visible in differences in for example length of reference lists, or age of cited literature), it is usually not meaningful to compare directly the raw citation impact of publications from one (sub)field with those of a different (sub)field. Therefore, in our studies raw citation impact scores are compared to the impact of similar publications within the same journal, or within the same (sub)field.

We start this final section of the introduction with a few general comments on the use of bibliometric indicators for the assessment of research performance. It is our experience in previous studies on research performance in the natural and life sciences, medicine, the humanities, and in the social and behavioral sciences, that bibliometric indicators provide useful information to a peer review committee evaluating research performance, or al least a group of experts. These studies revealed a fair correspondence between the results of bibliometric analyses on the one hand, and judgments on scientific quality by peers on the other hand. In our view, a quality judgment on a research unit, department or institute can only be given by peers, based on a detailed insight into content and nature of the research conducted by the group or institute in question. The citation-based indicators applied in this study, measure the impact at the short or middle-long term of research activities at the international research front, as reflected in publication and citation patterns. ***Impact and scientific quality are not necessarily identical concepts***.

Bibliometric indicators cannot be interpreted properly without background knowledge on both the research units that are evaluated, and the subfields in which the research units are active. In fact, in previous studies we have encountered a few cases in which a bibliometric indicator pointed in one direction (e.g., a low impact), while statements by peers or even other indicators pointed in another direction (e.g., a high quality). Analyzing such discrepancies from a bibliometric point of view, specific limitations related to the bibliometric methodology applied in the study in question may be identified. While in most cases such limitations do hardly affect the results or have no effect at all, in exceptional cases the bibliometric outcomes may provide an incomplete or even distorted picture. For instance, the classification of journals into subfields ('journal categories') may be less appropriate for some research units, particularly when they are active in topics of a multidisciplinary nature. Then, in the calculation of the impact compared to the world subfield citation average, this world average may not be representative for the subfield in which such a research group or institute is active. If there are strong indications that the definition of the (sub)field in terms of WOS journal categories is inadequate, then the journal-based world average is more appropriate. In particular, this latter case pertains to developing new interdisciplinary fields.

A second limitation concerns the coverage of the Citation Indices (CI). In specific subfields, particularly in applied or technical sciences, the WOS coverage may be less adequate. Consequently, for research units who are active in such technical/applied subfields, the bibliometric results may provide an incomplete picture. A second point concerns non-WOS publications (e.g., articles in journals that are not or no longer covered by CI, or book chapters and books). For a number of research units, valuable additional information may be obtained by retrieving impact data for non-WOS publications.

Another example of a limitation of bibliometric analysis relates to time delays. It may take several years for a collection of papers to generate a high impact. We have analyzed research units that had generated only a moderate impact at the time. Confronted with the bibliometric results, several peers stated that these research units had recently made important contributions to the field. When we updated the results after a few years, several research units indeed showed a sharply rising impact curve.

We do not wish to imply that all discrepancies between bibliometric indicators and peer judgments are necessarily due to problems or limitations of the bibliometric methods applied. Equally, it would not be appropriate to attribute such discrepancies only to peers expressing incorrect or biased views on the scientific quality of a research unit. Still reasoning from the point of view of the bibliometrician, discrepancies between bibliometric indicators and peer judgments often constitute a research problem in itself and often, a considerable effort is required to examine a discrepancy in sufficient detail.

Nevertheless, also peer review has its disadvantages. ***Therefore, the appropriate combination of peer-based qualitative assessment and quantitative, particularly bibliometric indicators appears to be the most successful approach in order to reinforce objectivity, transparency, comparability and reproducibility in the assessment of research performance.***

1.2 Datasets used in the study

The present study 'Research Impact of the Dutch University Medical Centers (UMCs)' is practically based on one central database or dataset, and two complimentary datasets. The central database is the Web of Science database, further named as WoS. This database is at CWTS turned from a bibliographic system to mainly collect publications, and do literature search, into a fully-fledged bibliometric system. This means that various steps are taken in to make this ready for bibliometric analysis, in which standardization plays an important role:1) additional datasets are created in order to perform analyses in which a certain degree of standardization or normalization can be applied, e.g., in citation impact analysis, 2) unification of information is applied, in order to create higher degrees of accuracy while analyzing performance, e.g., on the level of main institutions.

In a world in which research related information continuously grows, identifiers are extremely important. The WoS database has its' own identifiers (UT-codes), which make it possible to identify individual publications and all its characteristics. Unfortunately, these UT-codes are proprietary by nature, which means that these cannot be used to link to other datasets. In the study we used two complimentary datasets to WoS. The first we mention here is the Altmetrics.com dataset, which contains information on social media and related information. While Twitter and Facebook mentions form the core of the Altmetrics database, we here focus on other aspects of mentioning of scientific outputs. The focus will be on News items, on Policy documents, and on Clinical guidelines, three forms of communication through which we think we can measure societal relevance best, as in the mentioning a certain degree of engagement or commitment with the scientific output is involved, that might not be present while looking at Twitter and Facebook mentions (think about the reason why a certain publication from any of the Dutch UMCs will be mentioned in either a policy document, or a clinical guideline, we assume a genuine reason for that). Altmetric mentions function in a digital universe, so for linking up Altmetric pieces of information, we use the DOIs (Digital Object Identifiers) to link this information to WoS covered publications. This of course assumes presence of DOIs in both systems, and WoS has been substantially enriched over the last couple of years with DOIs. The second complimentary source we use in the study is the database that discloses openness on scientific publications, namely the Unpaywall database. The Unpaywall database has become the standard in the business quite quickly, and distinguishes between publications in closed or toll-access format, versus open access publications. Within this openness status, the system distinguishes various types of Open access publishing (Gold, Green, Hybrid, and Bronze).

1.3 Specifics on data collection

The present study relates to the publication output of the UMCs at Dutch universities. The UMCs supplied publication lists to CWTS, which were matched with the CWTS in-house bibliometric data-system. The bibliometric analysis is covering various periods of analysis: for the maps we used the publications from 2018 (thereby having 2019 as a potential year for citation analysis), for the analysis of scientific activity and collaboration we used the period 2013-2018/2019 , while we used the period 2004-2018/2019 for the trend analysis of citation impact development for all seven UMCs. This study is an update of the study conducted in 2017, for the update CWTS was supplied with the years 2018 and 2019 as additional publication years to the data collected for the previous studies.
We considered only papers classified in the WoS as normal articles and reviews, published in source serials processed for the WoS database. Please note that in the analysis letters are excluded. Other document types, such as meeting abstracts, 'editorials', 'editorial material', corrections, comments, and book reviews were also not included. Also, papers in non-WoS source journals are not counted. A few journals are only partially processed for the WoS. Here, only papers processed for the WoS were included.

2. Bibliometric indicators

2.1 Output and impact indicators

We calculated the following indicators. The numbering of the indicators corresponds to the position these indicators have in the data tables.
A *first* statistic gives the total number of papers published by the research unit during the entire period ***(P)***. We considered only papers classified as *normal articles* and *reviews*. Letters, meeting abstracts, corrections, and editorials are *not* included. In a few cases, a paper is published in a journal for which no citation data are available, or that is not assigned to a CI

journal category. These papers are not considered in the calculation of the indicators presented in the tables below.

The next indicator gives the total number of citations received, without self-citations (***TCS***). In the calculation of all our impact indicators, we disregard author self-citations. We classify a citation as an author self-citation if the citing publication and the cited publication have at least one author name (i.e., last name and initials) in common. In this way, we ensure that our indicators focus on measuring only the contribution and impact of the work of a researcher on the work of other members of the scientific community. Sometimes self-citations can serve as a mechanism for self-promotion rather than as a mechanism for indicating relevant related work. The impact of the work of a researcher on his own work is therefore ignored.

A next indicator is the average number of citations per publication calculated while self-citations are not included (***MCS)***.

Main scientific impact indicators

The overall field normalized impact indicator for an institute output is ***MNCS***, the Mean Normalized Citation Score. As this indicator focuses on the broader environment of the group's output, this indicator seems the most suitable indicator of the international position of a research unit. Our mean normalized citation score indicator, denoted by ***MNCS***, provides a more sophisticated alternative to the MCS indicator. The ***MNCS*** indicator is similar to the MCS indicator except that a normalization is being applied to correct for differences in citation characteristics between publications from different scientific fields and between publications of different ages (in the case of a variable-length citation window). To calculate the ***MNCS*** indicator for a unit, we first calculate the normalized citation score of each publication of the unit. The normalized citation score of a publication equals the ratio of the actual and the expected number of citations of the publication, where the expected number of citations is defined as the average number of citations of all publications in WoS belonging to the same field and having the same publication year. The field to which a publication belongs is determined by the micro-clusters to which the publication is attributed.

The ***MNCS*** indicator is obtained by averaging the normalized citation scores of all publications of a unit. If a unit has an ***MNCS*** indicator score of one, this means that on average the actual number of citations of the publications of the unit equals the expected number of citations. In other words, on average the publications of the unit have been cited equally frequently as publications that are similar in terms of field and publication year using the same citation window. An ***MNCS*** indicator score of, for instance, 2 means that on average the publications of a unit have been cited twice as frequently as would be expected based on their field and publication year.

A second important indicator, ***MNJS***, is above (below) 1.0 if the citation score of the journal set in which the research unit has published exceeds the citation score of all papers published in the subfield(s) to which the journals belong. In this case, one can conclude that the research unit publishes in journals with a relatively high (low) impact.

In addition to the ***MNCS*** indicator, we use another important impact indicator. This is the proportion of publications belonging to the top 10% most highly cited, denoted by PP(top 10%). For each publication of a research group, we determine whether it belongs to the top 10% based on its number of citations of all WoS publications in the same field (i.e., the micro-cluster) and from the same publication year. The ***PP(top 10%)*** indicator of a research entity equals the proportion of its publications belonging to this top 10%. If a research group has a ***PP(top 10%)*** indicator of 10%, it means that the actual number of top 10% publications of the group equals the expected number. A ***PP(top 10%)*** indicator of, for instance, 20% means that a group has twice as many top 10% publications as expected. Of course, the choice to focus on top 10% publications is somewhat arbitrary. Next to the ***PP(top 10%)*** indicator, we can also calculate ***PP(top 1%)***, ***PP(top 2%)***, ***PP(top 5%)***, or ***PP(top 20%)*** and ***PP(top 50%)*** indicators. In the main tables, we use the ***PP(top 10%)*** indicator. The other PP(top *x*%) indicators are presented as a separate analysis in the study.

In the Introduction we discussed the inadequacy of journal classification in the assessment of research units. The experience of using journal classification in bibliometrics for normalization purposes has led to the conclusion that basically these classifications do not function properly in general. An important element is the citation density one can observe even within the set of journals grouped under one category (high density versus low density areas, even within one journal subject category). Therefore, a system has been developed that distributes publications over clusters, each reflecting a research field or specialty. The composition of the clusters is based upon citation traffic between publications in the WoS, and citation density determines how a publication is clustered. At the lowest level of granularity we observe some 4000 clusters (in comparison to some 250 journal categories), which means that the level of accuracy of what belongs together is much more detailed compared to journal categories. A paper can only by grouped in one cluster.

2.2 Explanation of the citation impact measurement

In the standard tables, we apply the method in which citation impact is measured for five year maximum, in a block of publication years of four year maximum. This works as follows: for the first year in a four year block the impact is measured for five years, for the next year in the block we apply a four year citation window, for the third year in the block we apply a three year citation window, and for the last year in the block we apply a two year citation window. As an example, for the publication years 2015-2018, we apply a citation window that stretches the period 2015-2019, with a five year citation window 2015-2019 for the papers of 2015, a four year citation window of 2015-2019 for the 2016 publications, a three year citation window (20172019) for the year 2017, and finally a two year citation window (2018-2019) for the 2018 publications. This moves through time like roof tiles, in which the next period overlaps the previous. This approach has several advantages, namely in the first place the full usage of all publication years in the analysis in a similar fashion, which creates a consistent approach, and secondly, the aspect that publications contribute to each block in a different way, maturing in time, and overall creating a more smooth development of research impact measurements. With respect to this latter aspect, we have to stress that due to smaller output numbers, on lower levels of aggregation (such as projects or small teams), these outcomes tend to fluctuate more as compared with output numbers related to aggregates on higher levels (universities, UMCs, or divisions within UMCs).

2.3 Analysis of scientific collaboration

The analysis of the various types of scientific cooperation is based upon a typology of papers, which is based on the addresses attached to the publications. In case of the paper carrying only one address, the publications is automatically labeled as a single institute publication. In case of the appearance of at least two different country names on one publication, the publication is automatically considered an international cooperation. The remaining set of publications, carrying two or more addresses within one country, are considered to be the result of national cooperation.

Any classification such as this one has some drawbacks. For example, the typology applied has the disadvantage that in the case of international cooperation publications, if a paper also carries two addresses from one country, the international dimension is the dominant factor in labeling the publication. Furthermore, in case of publications labeled as national cooperation, it can happen that these are actually two addresses of one and the same main institution, which makes it an intra-mural cooperation. However, the typology has been designed in order to have mutually exclusive classes thus simplifying the analysis of collaboration networks (the strength of such typology is clearly visible through an analysis of international scientific cooperation links, see van Leeuwen, 2009).

2.4 Maps of science

In order to get an impression of the composition of the total field of biomedical research in the Dutch UMCs, we created so-called maps of science. The methodology used for this developed by CWTS (www.vosviewer.com). This analysis is based upon the titles and abstracts of publications of the publications of Dutch UMCs in the year 2018, as far as covered in the WoS. In this report only publications in clusters with over 15 publications in one year were selected, and clusters with a joint mean normalized citation impact score of higher than 1.5. The publication output is semantically parsed, during which meaningful noun phrases are identified. Next, these noun phrases are analyzed through a cluster algorithm on their cooccurrences within the titles and abstracts per publication, aggregated over the total output. This leads to a map with a certain structure, and particular specific characteristics, such as:

- Distance between words indicate relatedness, the closer the words, the stronger their relation, and vice versa;
- The font size of the terms indicate the frequency of occurrences;
- The color indicates a stronger local relationship.

What we did next was the creation on top of the overall map, the so-called overlay maps, which indicate the specific activities of each of the seven UMCs involved. The color coding in the overlay maps show the specific focus of each of the units in the study. The maps are visible as screen shots in the report, while the methodology is actually an interactive tool, that allows for varying the structure created. The interactive tool with data from the UMCs will be accessible from the NFU website.

With respect to the interpretation of the general map and the overlay maps we should mention that the maps display a multi-dimensional structure in a two-dimensional space, which, means that sometimes the distance between words is not directly readable as relatedness.